Statistica Sinica **6**(1996), 233-244

# TRIMMED ESTIMATION IN THE ERRORS-IN-VARIABLES MODEL

Longcheen Huwang and Stanley Li

*National Tsing Hua University*

*Abstract.* We consider a method of defining trimmed estimators of coefficients in an errors-in-variables model using the trimmed least squares estimators suggested by Koenker and Bassett (1978). The resultant estimators are consistent and asymptotically normal. In terms of the asymptotic relative efficiency, these trimmed estimators are more efficient than the traditional ones when the regression error in the errors-in-variables model has a heavy tailed distribution. A lower bound for the asymptotic relative efficiency is also established under some assumptions.

Key words and phrases: Errors-in-variables, trimmed least squares estimator, asymptotic relative efficiency.

## 1. Introduction

Errors-in-variables models arise from the study of regression models wherein the covariate is measured with error. Suppose that there are unobservable "true" random variables $(u_i, \eta_i)$ which satisfy a linear relation,

$$\eta_i = \alpha + \beta u_i. \tag{1.1}$$

However, we can only observe $(X_i, Y_i)$ which are the true random variables plus additive measurement errors $(\delta_i, \varepsilon_i)$, i.e.

$$X_i = u_i + \delta_i, \quad Y_i = \eta_i + \varepsilon_i, \quad i = 1, \ldots, n. \tag{1.2}$$

It is assumed that $\{u_i\}$, $\{\delta_i\}$, and $\{\varepsilon_i\}$ are three i.i.d. sequences of random variables, and that $u_i \sim N(m_u, \sigma_u^2)$, $\delta_i \sim N(0, \sigma_\delta^2)$, and $\varepsilon_i$ has a continuous distribution function $F$ which is symmetric about 0. Moreover, it is also assumed that $F$ has a continuous density $f$ that is positive on the support of $F$. Model (1.1)-(1.2) is called a structural errors-in-variables model since the $u_i$'s are assumed to be i.i.d. random variables. This is different from a functional errors-in-variables model where the $u_i$'s are assumed to be unknown constants, not a random sample. Errors-in-variables model arises in many applications. Surveys of results can be found in Moran (1971), Kendall and Stuart (1979), Anderson (1984), Fuller (1987), and Cheng and Van Ness (1991).

It is well known that without extra assumptions about the parameters, model (1.1)-(1.2) is unidentifiable. For example, parameter sets $(\alpha, \beta, m_u, \sigma_u^2, \sigma_\delta^2, F) = (0, 1, 1, 1, 1, F)$ and $(1 - 1/\sqrt{2}, 1/\sqrt{2}, 1, 2, 0, F)$ are both such that $(X_i, Y_i)$ have the same distribution. The present paper assumes that the error variance $\sigma_\delta^2$ is known. In fact, knowledge of the error variance $\sigma_\delta^2$ is realistic. For example, the $X_i$'s may arise as the averages of repeated measurements on the $u_i$'s in cases where no corresponding repeated measurements on the $\eta_i$'s are available, i.e. "partial replications". These repeated measurements can be used to form an independent estimator of $\sigma_\delta^2$ (Madansky (1959)).

The traditional estimators of $\alpha$ and $\beta$, the unknown key parameters, are defined by

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}, \qquad \hat{\beta} = \frac{S_{XY}}{S_X^2 - \sigma_\delta^2}, \tag{1.3}$$

where $\bar{X} = \sum X_i/n$, $\bar{Y} = \sum Y_i/n$, $S_X^2 = \sum(X_i - \bar{X})^2/n$, $S_Y^2 = \sum(Y_i - \bar{Y})^2/n$, and $S_{XY} = \sum(X_i - \bar{X})(Y_i - \bar{Y})/n$. Note that under the assumption that $\varepsilon_i$ in (1.2) is normally distributed, these $\hat{\alpha}$ and $\hat{\beta}$ are also the maximum likelihood estimators of $\alpha$ and $\beta$ whenever $S_Y^2(S_X^2 - \sigma_\delta^2) - S_{XY}^2 > 0$ (Fuller (1987, p.14) and Kendall and Stuart (1979, p.405)). Although these estimators are consistent and asymptotically normal, they are, however, inefficient when $F$ has heavier tails than the Gaussian distribution. Moreover, these estimators also possess high sensitivity to spurious data $Y_i$'s. The presence of spurious observations $Y_i$'s can be modeled by letting $F$ be a mixture of the distribution of the "good" data, for instance, standard normal, and that of the "bad" data, for instance, normal with variance exceeding 1. Such a distribution $F$ will have heavier tails than a normal distribution. For the location model, three classes, M, L, and R of estimators have been suggested as alternatives to the traditional sample mean (see Lehmann (1983) for an introduction). Among the L estimators, the trimmed mean is particularly attractive since it is efficient and easy to compute under most circumstances. Stigler (1977, p.1070) applied robust estimators to data from 18th- and 19th-century experiments design to measure basic physical constants. He concluded that the 10% trimmed mean is preferable as a recommended estimator.

Consider the linear model

$$\boldsymbol{W} = D\boldsymbol{\gamma} + \boldsymbol{Z}, \tag{1.4}$$

where $\boldsymbol{W} = (W_1, \ldots, W_n)'$, $D$ is a $n \times p$ matrix of known constants whose $i$th row is $\boldsymbol{d}_i'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)'$ is a vector of unknown parameters, and $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ is a vector of i.i.d. random variables with unknown distribution $G$. Koenker and Bassett (1978), who extended the concept of quantiles to the linear model, proposed a method of defining a regression analog to the trimmed mean. Let

$0 < \rho < 1$. They defined the $\rho$th regression quantile, denoted by $\gamma(\rho)$, to be any solution to the minimization problem:

$$\min_{\boldsymbol{b}\in R^p}\{\sum_{i\in A}(1-\rho)|W_i - \boldsymbol{d}_i'\boldsymbol{b}| + \sum_{i\in A^c}\rho|W_i - \boldsymbol{d}_i'\boldsymbol{b}|\},\ \ A = \{i : W_i - \boldsymbol{d}_i'\boldsymbol{b} < 0\}. \quad (1.5)$$

They showed that the regression quantiles have asymptotic behavior similar to those of sample quantiles in the one-sample problem. As Koenker and Bassett pointed out, regression quantiles can be computed by standard linear programming techniques (Meketon (1986)). They also suggested the following trimmed least squares estimator $\hat{\gamma}(\theta)$: Where $0 < \theta < 1/2$ is the trimmed proportion, remove from the sample any observations satisfying

$$W_i - \boldsymbol{d}_i'\gamma(\theta) \le 0 \qquad \text{or} \qquad W_i - \boldsymbol{d}_i'\gamma(1-\theta) \ge 0, \quad (1.6)$$

and calculate the least squares estimator using the remaining observations. Ruppert and Carroll (1980) made the following assumptions:

1. $G$ has a continuous density $g$ that is positive on the support of $G$.
2. Letting $(d_{i1},\ldots,d_{ip}) = \boldsymbol{d}_i'$ be the $i$th row of $D$, $d_{i1} = 1$ for $1 \le i \le n$ and $\sum_{i=1}^n d_{ij} = 0$ for $2 \le j \le p$.
3. $\lim_{n\to\infty}\max_{1\le i\le n, 1\le j\le p}|d_{ij}|/\sqrt{n} = 0$.
4. There exists positive definite $Q$ such that $\lim_{n\to\infty} D'D/n = Q$.

They showed that the estimator $\hat{\gamma}(\theta)$ satisfies

$$\sqrt{n}[\hat{\gamma}(\theta) - \gamma] = Q^{-1}n^{-1/2}\left\{\sum_{i=1}^n \boldsymbol{d}_i[\phi(Z_i) - E\phi(Z_i)] + \boldsymbol{\delta}(\theta)\right\} + o_p(1) \quad (1.7)$$

and

$$\sqrt{n}[\hat{\gamma}(\theta) - \gamma - \boldsymbol{\delta}(\theta)]\xrightarrow{L} N_p[\mathbf{0}, \sigma^2(\theta,G)Q^{-1}], \quad (1.8)$$

where

$$\phi(s) = \begin{cases} \eta_1/(1-2\theta), & \text{if } s < \eta_1, \\ s/(1-2\theta), & \text{if } \eta_1 \le s \le \eta_2, \\ \eta_2/(1-2\theta), & \text{if } \eta_2 < s, \end{cases}$$

$$\delta(\theta) = \frac{1}{1-2\theta}\int_{\eta_1}^{\eta_2} s dG(s), \quad \boldsymbol{\delta}(\theta) = (\delta(\theta), 0, \ldots, 0)',$$

$$\sigma^2(\theta,G) = \frac{1}{(1-2\theta)^2}\Big\{\int_{\eta_1}^{\eta_2}[s-\delta(\theta)]^2 dG(s) + \theta(\kappa_1^2+\kappa_2^2) - [\theta(\kappa_1+\kappa_2)]^2\Big\},$$

$$\eta_1 = G^{-1}(\theta), \quad \eta_2 = G^{-1}(1-\theta), \text{ and } \kappa_j = [\eta_j - \delta(\theta)]. \quad (1.9)$$

By comparing asymptotic variances, they also concluded that the trimmed least squares estimator turns out to be more efficient than the usual least squares estimator when $G$ is within the family of contaminated normal distributions and the extent of contamination is getting large. These contaminated normal distributions have long been used to study the behavior of statistical procedures under heavy-tailed distributions. They have the form

$$(1-\lambda)\Phi(s)+\lambda\Phi(\frac{s}{b}), \tag{1.10}$$

where $0<\lambda<1$ and $\Phi$ is the standard normal distribution function. Typically, $b>1$ and $\Phi(s/b)$ is the distribution of the "bad" data, whereas $\lambda$ is the proportion of "bad" data.

In this paper, we write errors-in-variables model (1.1)-(1.2) as a form of linear model (1.4). By the relations between the parameters of these two models, we define the trimmed estimators of $\alpha$ and $\beta$ in errors-in-variables model (1.1)-(1.2) through the trimmed least squares estimators for linear model (1.4). As will be shown in this work, these trimmed estimators are not only asymptotically normal but also more efficient than the traditional ones (estimators in (1.3)) when the distribution $F$ of the regression error $\varepsilon_i$ in model (1.1)-(1.2) has heavier tails than the Gaussian distribution.

In Section 2 we define the trimmed estimators for errors-in-variables model (1.1)-(1.2). The asymptotic distributions of these trimmed estimators are derived in Section 3. In Section 4 we compare the efficiencies of the trimmed estimators and the traditional ones. A lower bound for these efficiencies is also established under certain assumptions.

## 2. Trimmed Estimators

For errors-in-variables model (1.1)-(1.2), the degree to which the covariate $u_i$ has been contaminated with error can be measured by the ratio $r=\sigma_u^2/(\sigma_u^2+\sigma_\delta^2)$, called the reliability of $X_i$. Conditioning on $X_i$, $u_i$ is normally distributed with mean $(1-r)m_u+rX_i$ and variance $r\sigma_\delta^2$. Therefore, the second equation of (1.2) can be written as

$$Y_i=\alpha+\beta E(u_i|X_i)+\beta[u_i-E(u_i|X_i)]+\varepsilon_i=\alpha^*+\beta^*(X_i-\bar{X})+\varepsilon_i^*, \tag{2.1}$$

where

$$\alpha^*=\alpha+[r\bar{X}+(1-r)m_u]\beta,\ \beta^*=r\beta,\ \varepsilon_i^*=v_i+\varepsilon_i,\ v_i=\beta[u_i-rX_i-(1-r)m_u]. \tag{2.2}$$

Since $v_i$ is distributed as $N(0,\beta^2r\sigma_\delta^2)$ which is independent of $X_i$, consequently $\varepsilon_i^*$ and $X_i$ are independent. Thus, model (2.1) is a linear regression model. Let $0<\theta<1/2$ be the trimmed proportion and let $\hat{\alpha}^*(\theta)$ and $\hat{\beta}^*(\theta)$ be the trimmed least squares estimators of $\alpha^*$ and $\beta^*$ in (2.1) suggested by Koenker and Bassett

(1978) as described in Section 1. Then we define the trimmed estimators of $\alpha$ and $\beta$ in model (1.1)-(1.2) by

$$\hat{\alpha}(\theta) = \hat{\alpha}^*(\theta) - [\bar{X} + (\frac{1}{\hat{r}} - 1)\hat{m}_u]\hat{\beta}^*(\theta) = \hat{\alpha}^*(\theta) - \frac{1}{\hat{r}}\bar{X}\hat{\beta}^*(\theta), \;\; \hat{\beta}(\theta) = \frac{\hat{\beta}^*(\theta)}{\hat{r}}, \quad (2.3)$$

where $\hat{m}_u = \bar{X}$ and $\hat{r} = (S_X^2 - \sigma_\delta^2)/S_X^2$ are consistent estimators of $m_u$ and $r$, respectively. Note that unlike $\hat{\alpha}^*(\theta)$ and $\hat{\beta}^*(\theta)$ we do not remove any observations in calculating $\hat{m}_u$ and $\hat{r}$. Since the $X_i$'s are *i.i.d.* normal random variables, the estimators $\hat{m}_u$ and $\hat{r}$ without the trim procedure are more efficient than the trimmed ones.

**Comment**. The main reason for confining our approach to the structural model ($u_i$-random) is to take advantage of conditional expectation for $u_i$ given $X_i$. This is not defined for the functional model, where the $u_i$ are unknown fixed constants. However, under the assumptions that the limits $u^* = \lim_{n\to\infty} \bar{u}_n = \lim_{n\to\infty} \sum u_i/n$ and $v^* = \lim_{n\to\infty} \sum(u_i - \bar{u}_n)^2/n$ exist, it is not difficult to define the parallel trimmed estimators for the functional model that result in asymptotic consistency. Nevertheless, the asymptotic distributions of these estimators could be hard to establish.

## 3. Asymptotic Results

Although model (2.1) has a form as model (1.4), they differ in the nature of design matrix. The design matrix of model (1.4) is nonstochastic, whereas that of model (2.1) is stochastic. In order to apply the results of (1.7) and (1.8) to model (2.1), we need to show that the distribution $H$ of $\varepsilon_i^*$ in (2.1) satisfies the first assumption of Ruppert and Carroll (1980) as stated in Section 1, and that the $X_i$'s satisfy the assumptions 2, 3, and 4 of Ruppert and Carroll with probability 1.

From (2.2) we have $\varepsilon_i^* = v_i + \varepsilon_i$, where $v_i = \beta[u_i - rX_i - (1-r)m_u] \sim N(0, \beta^2 r \sigma_\delta^2)$. Since the distribution $F$ of $\varepsilon_i$ has a continuous density $f$ that is positive on the support of $F$, obviously the distribution $H$ of $\varepsilon_i^*$ satisfies the first assumption of Ruppert and Carroll. Also, from the structure of model (2.1), the $X_i$'s clearly satisfy assumption 2 with probability 1. For assumption 3, it suffices to show that

$$P\Big(\lim_{n\to\infty} \max_{1\le i\le n} \frac{(X_i - \bar{X})^2}{n} = 0\Big) = 1.$$

For every $\epsilon > 0$, we have

$$P\Big(\max_{1\le i\le n} \frac{(X_i - \bar{X})^2}{n} \le \epsilon\Big) = P\Big(\bigcap_{i=1}^{n} [\frac{(X_i - \bar{X})^2}{n} \le \epsilon]\Big)$$

$$\geq 1-\sum_{i=1}^{n} P\Big(\frac{(X_i-\bar{X})^2}{n}>\epsilon\Big)>1-\frac{1}{n^2\epsilon^3}E(X_i-\bar{X})^6.$$

Consequently,

$$\sum_{i=1}^{\infty} P\Big(\max_{1\leq i\leq n}\frac{(X_i-\bar{X})^2}{n}>\epsilon\Big)<\frac{1}{\epsilon^3}E(X_i-\bar{X})^6\sum_{i=1}^{\infty}\frac{1}{n^2}<\infty$$

due to $E(X_i-\bar{X})^6<\infty$. Applying the convergent part of Borel-Cantelli lemma, we have

$$P\Big(\limsup_{n\to\infty}\max_{1\leq i\leq n}\frac{(X_i-\bar{X})^2}{n}>\epsilon\Big)=0.$$

Therefore,

$$\lim_{n\to\infty}\max_{1\leq i\leq n}\frac{(X_i-\bar{X})^2}{n}=0 \ \text{ a.s.}$$

For assumption 4, since $S_X^2\to\sigma_X^2$ $(=\sigma_u^2+\sigma_\delta^2)$ a.s., we have

$$\frac{1}{n}\begin{pmatrix}1 & \cdots & 1\\ X_1-\bar{X} & \cdots & X_n-\bar{X}\end{pmatrix}\begin{pmatrix}1 & \cdots & 1\\ X_1-\bar{X} & \cdots & X_n-\bar{X}\end{pmatrix}'$$
$$=\begin{pmatrix}1 & 0\\ 0 & S_X^2\end{pmatrix}\to\begin{pmatrix}1 & 0\\ 0 & \sigma_X^2\end{pmatrix} \ \text{ a.s.}$$

**Lemma 1.** *Let model (1.1)-(1.2) hold with $\sigma_\delta^2$ known, and let $\hat{\alpha}^*(\theta)$ and $\hat{\beta}^*(\theta)$ be the trimmed least squares estimators of $\alpha^*$ and $\beta^*$ in (2.1) with trimmed proportion $0<\theta<1/2$. Then*

$$\sqrt{n}[\hat{\alpha}^*(\theta)-\alpha^*,\,\hat{\beta}^*(\theta)-\beta^*]'\xrightarrow{L}N_2[\mathbf{0},\sigma^2(\theta,H)\Sigma^{-1}],\tag{3.1}$$

*where*

$$\sigma^2(\theta,H)=\frac{1}{(1-2\theta)^2}\Big[\int_{\eta_1}^{\eta_2}s^2dH(s)+\theta(\eta_1^2+\eta_2^2)\Big]=\frac{2}{(1-2\theta)^2}\Big[\int_{0}^{\eta_2}s^2dH(s)+\theta\eta_2^2\Big],$$

$$\eta_1=H^{-1}(\theta),\quad \eta_2=H^{-1}(1-\theta),\quad \Sigma=\begin{pmatrix}1 & 0\\ 0 & \sigma_X^2\end{pmatrix},\tag{3.2}$$

*and $H$ is the distribution function of $\varepsilon_i^*$ in (2.1).*

**Proof.** Since $v_i\sim N(0,\beta^2r\sigma_\delta^2)$ and $\varepsilon_i$ are independent and $\varepsilon_i$ has a continuous distribution $F$ which is symmetric about 0, the distribution $H$ of $\varepsilon_i^*=v_i+\varepsilon_i$ is continuous and symmetric about 0 as well. Therefore, $\delta(\theta)=0$ in (1.9) because $\eta_1=H^{-1}(\theta)=-H^{-1}(1-\theta)=-\eta_2$. Now conditioning on $X_i, 1\leq i\leq n$, and applying (1.8), we have

$$\sqrt{n}[\hat{\alpha}^*(\theta)-\alpha^*,\,\hat{\beta}^*(\theta)-\beta^*]'\xrightarrow{L}N_2[\mathbf{0},\sigma^2(\theta,H)\Sigma^{-1}].$$

Note that both $\sigma^2(\theta, H)$ and $\Sigma$ above do not depend on $X_i, 1 \le i \le n$. By the bounded convergence theorem, the limiting result holds unconditionally.

From (2.2) and (2.3), we have

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 & -[\bar{X} + (1/r - 1)m_u] \\ 0 & 1/r \end{pmatrix} \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix}$$

and

$$\begin{pmatrix} \hat{\alpha}(\theta) \\ \hat{\beta}(\theta) \end{pmatrix} = \begin{pmatrix} 1 & -\bar{X}/\hat{r} \\ 0 & 1/\hat{r} \end{pmatrix} \begin{pmatrix} \hat{\alpha}^*(\theta) \\ \hat{\beta}^*(\theta) \end{pmatrix}.$$

Therefore,

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}(\theta) - \alpha \\ \hat{\beta}(\theta) - \beta \end{pmatrix} = \sqrt{n} \begin{pmatrix} 1 & -\bar{X}/\hat{r} \\ 0 & 1/\hat{r} \end{pmatrix} \begin{pmatrix} \hat{\alpha}^*(\theta) - \alpha^* \\ \hat{\beta}^*(\theta) - \beta^* \end{pmatrix} + \sqrt{n} \begin{pmatrix} 0 & (1 - 1/\hat{r})\bar{X} + (1/r - 1)m_u \\ 0 & 1/\hat{r} - 1/r \end{pmatrix} \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix}. \quad (3.3)$$

Using this relation and Lemma 1, we can prove the main theorem in this section.

**Theorem 1.** *Let model* (1.1) – (1.2) *hold with* $\sigma_\delta^2$ *known, and let* $\hat{\alpha}(\theta)$ *and* $\hat{\beta}(\theta)$, *which are defined in* (2.3), *be the trimmed estimators of* $\alpha$ *and* $\beta$ *in this model with trimmed proportion* $0 < \theta < 1/2$. *Then*

$$\sqrt{n} \left( \hat{\alpha}(\theta) - \alpha,\, \hat{\beta}(\theta) - \beta \right)' \xrightarrow{L} N_2(\mathbf{0}, \Gamma),$$

*where*

$$\Gamma = \begin{pmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix}, \quad \sigma_\alpha^2 = \sigma^2(\theta, H)(1 + \frac{m_u^2}{r^2\sigma_X^2}) + \beta^2(1-r)^2(\sigma_X^2 + \frac{2m_u^2}{r^2}),$$

$$\sigma_\beta^2 = \frac{1}{r^2}\left[2\beta^2(1-r)^2 + \frac{\sigma^2(\theta,H)}{\sigma_X^2}\right], \text{ and } \sigma_{\alpha\beta} = \frac{-m_u}{r^2}\left[2\beta^2(1-r)^2 + \frac{\sigma^2(\theta,H)}{\sigma_X^2}\right]. \quad (3.4)$$

**Proof.** Applying a Taylor expansion, we can show that

$$\begin{pmatrix} 1 & -\bar{X}/\hat{r} \\ 0 & 1/\hat{r} \end{pmatrix} = \begin{pmatrix} 1 & -m_u/r \\ 0 & 1/r \end{pmatrix} + O_p(\frac{1}{\sqrt{n}}). \quad (3.5)$$

By Lemma 1, (3.5), and Slutsky's theorem, the first term on the right hand side of Equation (3.3) converges in distribution to $N_2(\mathbf{0}, \Gamma_1)$, where

$$\Gamma_1 = \sigma^2(\theta, H) \begin{pmatrix} 1 + m_u^2/(r^2\sigma_X^2) & -m_u/(r^2\sigma_X^2) \\ -m_u/(r^2\sigma_X^2) & 1/(r^2\sigma_X^2) \end{pmatrix}. \quad (3.6)$$

Conditioning on $X_i$, $1 \le i \le n$, and by (1.7) we have

$$\sqrt{n}[\hat{\alpha}^*(\theta) - \alpha^*, \hat{\beta}^*(\theta) - \beta^*]' = \Sigma^{-1} n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_i - \bar{X} \end{pmatrix} [\phi(\varepsilon_i^*) - E\phi(\varepsilon_i^*)] + o_p(1)$$

$$= n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} \phi(\varepsilon_i^*) - E\phi(\varepsilon_i^*) \\ ((X_i - \bar{X})/S_X)[\phi(\varepsilon_i^*) - E\phi(\varepsilon_i^*)]/\sigma_X \end{pmatrix} + o_p(1), \quad (3.7)$$

where $\phi$ is defined by (1.9). It follows that (3.7) holds unconditionally by the bounded convergence theorem. Since $(\bar{X}, S_X^2)$ is sufficient and complete for $(m_u, \sigma_X^2)$, the ancillary statistic $(X_i - \bar{X})/S_X$ is independent of $(\bar{X}, S_X^2)$. Therefore, from (3.5) and (3.7) we conclude that the two terms on the right hand side of Equation (3.3) are asymptotically independent.

Again, applying a Taylor expansion the second term on the right hand side of (3.3) can be written as

$$\sqrt{n} \begin{pmatrix} (1 - 1/r)(\bar{X} - m_u) - (1/\hat{r} - 1/r) m_u \\ 1/\hat{r} - 1/r \end{pmatrix} r\beta + o_p(1). \quad (3.8)$$

Consequently, it converges in distribution to $N_2(\mathbf{0}, \Gamma_2)$, where

$$\Gamma_2 = \begin{pmatrix} \beta^2(1-r)^2(\sigma_X^2 + 2m_u^2/r^2) & -2\beta^2 m_u (1-r)^2/r^2 \\ -2\beta^2 m_u (1-r)^2/r^2 & 2\beta^2(1-r)^2/r^2 \end{pmatrix}. \quad (3.9)$$

Now the result follows from (3.6) and (3.9).

**Remark 1.** Let $\hat{\alpha}$ and $\hat{\beta}$, as defined in (1.3), be the traditional consistent estimators of $\alpha$ and $\beta$ in model (1.1)-(1.2) (which are also the maximum likelihood estimators under some extra conditions). Then by a similar proof of Theorem 1, we have

$$\sqrt{n}\left(\hat{\alpha} - \alpha, \hat{\beta} - \beta\right)' \xrightarrow{L} N_2(\mathbf{0}, \Gamma'),$$

where $\Gamma'$ is the same as $\Gamma$ in (3.4) except that $\sigma^2(\theta, H)$ is replaced by $\sigma_{\varepsilon^*}^2$, the variance of $\varepsilon_i^*$ in (2.1), applied to $\Gamma'$.

## 4. Comparison of Efficiency

In this section we take the distribution function $F$ of $\varepsilon_i$ in model (1.1)-(1.2) to be a member in the family of contaminated normal distributions defined by (1.10). Although these distributions have heavier tails than a normal distribution, the heavy tails are not as serious as that of the Cauchy distribution. Hence they are more likely to be used in practical situations. Since $v_i$ and $\varepsilon_i$ are independent and $v_i \sim N(0, \beta^2 r \sigma_\delta^2)$, the distribution function $H$ of $\varepsilon_i^* (= v_i + \varepsilon_i)$ in model (2.1) is the convolution of $\Phi(s/\sqrt{\beta^2 r \sigma_\delta^2})$ and $(1-\lambda)\Phi(s) + \lambda\Phi(s/b)$. Consequently,

$$H(s) = (1-\lambda)\Phi\left(\frac{s}{\sqrt{1 + \beta^2 r \sigma_\delta^2}}\right) + \lambda\Phi\left(\frac{s}{\sqrt{b^2 + \beta^2 r \sigma_\delta^2}}\right). \quad (4.1)$$

From this and (3.2), we have the variance of $\varepsilon_i^*$,

$$\sigma_{\varepsilon^*}^2 = (1-\lambda)(1+\beta^2 r\sigma_\delta^2) + \lambda(b^2+\beta^2 r\sigma_\delta^2) = (1-\lambda+\lambda b^2) + \beta^2 r\sigma_\delta^2, \tag{4.2}$$

and

$$\sigma^2(\theta,H) = \frac{2}{(1-2\theta)^2}\left[(1-\lambda)\int_0^{\eta_2} s^2 d\Phi\left(\frac{s}{\sqrt{1+\beta^2 r\sigma_\delta^2}}\right) + \lambda\int_0^{\eta_2} s^2 d\Phi\left(\frac{s}{\sqrt{b^2+\beta^2 r\sigma_\delta^2}}\right) + \theta\eta_2^2\right]. \tag{4.3}$$

Suppose that $\{\delta_{in}\}, i=1,2$, are two sequences of estimators of the parameter $g(\gamma)$ based on the $n$ observations $(X_i,Y_i), 1\le i\le n$, such that

$$\sqrt{n}[\delta_{in}-g(\gamma)] \xrightarrow{L} N[0,\tau_i^2(\gamma)], \tau_i^2(\gamma)>0.$$

Then the asymptotic relative efficiency (ARE) of $\{\delta_{1n}\}$ with respect to $\{\delta_{2n}\}$ is defined by $e_{1,2}=\tau_2^2(\gamma)/\tau_1^2(\gamma)$.

In Table 1 we tabulate the ARE's of $\hat\beta(\theta)$ with respect to $\hat\beta$ for $\beta=1$, $m_u=1$, $\sigma_u^2=1$, and a few choices of $\sigma_\delta^2$, $\lambda$, $b$, and $\theta$ (the results of those of $\hat\alpha(\theta)$ with respect to $\hat\alpha$ are quite similar and hence are not reported here). Several conclusions can be drawn from this table.

1. For $\lambda=0$ (this corresponds to $\varepsilon_i$ as well as $\varepsilon_i^*$ being normally distributed), the ARE's are all less than 1. However, when the trimmed proportions are less than 20%, these ARE's do not fall too low. For $\lambda=0.1$ and 0.25, the ARE's are considerably greater than 1 under almost all circumstances. Only in some rare situations ( for example, $\lambda=0.1, r=0.9, b=3$, and $\theta=50\%$) the ARE's are little less than 1. This poor efficiency is the consequence of over-trimming.

2. With other parameters being fixed, the ARE is increasing in $b$. This accords with the intuition that $\hat\beta(\theta)$ turns to be more efficient when the heavy-tailed phenomenon becomes serious. Although for $\lambda=0.1$ and 0.25 the ARE seems to be increasing in $r$ (or decreasing in $\sigma_\delta^2$) as other parameters are fixed, this is not necessarily the case (see, for example, $\lambda=0$). There is a trade-off among the parameters.

3. As a whole, the $10\%-20\%$ trimmed estimators $\hat\beta(\theta)$ provide much better protection than $\hat\beta$ against heavy contamination ($\lambda$ large or b large), while at the same time giving up little efficiency in the normal case ($\lambda=0$). They emerge as the recommended estimators.

From Table 1, it is seen that the ARE's of $\hat\beta(\theta)$ with respect to $\hat\beta$ are uniformly high for all $\theta$'s, and this raises the question of how these ARE's can fall if the distribution function $F$ of $\varepsilon_i$ in (1.2) does not belong to the family of contaminated normal distributions. The answer is given by the following theorem.

Table 1. ARE of $\hat{\beta}(\theta)$ with respect to $\hat{\beta}$
$\beta = 1, m_u = 1, \sigma_u^2 = 1$

| | | | ARE | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\lambda$ | r | b | $\theta$ 2.5% | 5% | 10% | 15% | 20% | 30% | 50% |
| 0 | 1.0 | | 0.99 | 0.97 | 0.94 | 0.91 | 0.87 | 0.80 | 0.64 |
| | 0.9 | | 0.99 | 0.97 | 0.94 | 0.91 | 0.88 | 0.80 | 0.64 |
| | 0.8 | | 0.98 | 0.98 | 0.95 | 0.92 | 0.88 | 0.81 | 0.66 |
| | 0.7 | | 0.99 | 0.97 | 0.95 | 0.92 | 0.89 | 0.83 | 0.68 |
| | 0.6 | | 0.99 | 0.98 | 0.96 | 0.93 | 0.91 | 0.85 | 0.71 |
| 0.1 | 1.0 | 3 | 1.29 | 1.36 | 1.38 | 1.36 | 1.33 | 1.24 | 1.00 |
| | | 5 | 1.84 | 2.31 | 2.46 | 2.46 | 2.43 | 2.27 | 1.83 |
| | | 10 | 2.51 | 6.57 | 7.52 | 7.62 | 7.52 | 7.08 | 5.74 |
| | 0.9 | 3 | 1.25 | 1.31 | 1.33 | 1.31 | 1.28 | 1.19 | 0.96 |
| | | 5 | 1.78 | 2.17 | 2.30 | 2.30 | 2.25 | 2.11 | 1.71 |
| | | 10 | 2.47 | 6.02 | 6.79 | 6.92 | 6.84 | 6.44 | 5.24 |
| | 0.8 | 3 | 1.22 | 1.27 | 1.28 | 1.27 | 1.23 | 1.15 | 0.93 |
| | | 5 | 1.71 | 2.02 | 2.13 | 2.13 | 2.08 | 1.96 | 1.59 |
| | | 10 | 2.41 | 5.44 | 6.09 | 6.19 | 6.09 | 5.77 | 4.73 |
| | 0.7 | 3 | 1.19 | 1.22 | 1.23 | 1.21 | 1.19 | 1.11 | 0.91 |
| | | 5 | 1.62 | 1.87 | 1.95 | 1.94 | 1.91 | 1.80 | 1.49 |
| | | 10 | 2.33 | 4.84 | 5.36 | 5.44 | 5.36 | 5.08 | 4.22 |
| | 0.6 | 3 | 1.15 | 1.18 | 1.18 | 1.17 | 1.14 | 1.07 | 0.90 |
| | | 5 | 1.52 | 1.72 | 1.78 | 1.77 | 1.74 | 1.66 | 1.39 |
| | | 10 | 2.24 | 4.24 | 4.63 | 4.69 | 4.63 | 4.41 | 3.72 |
| 0.25 | 1.0 | 3 | 1.19 | 1.40 | 1.62 | 1.69 | 1.69 | 1.61 | 1.33 |
| | | 5 | 1.26 | 1.70 | 2.93 | 3.35 | 3.47 | 3.43 | 2.86 |
| | | 10 | 1.32 | 1.88 | 6.94 | 10.51 | 11.44 | 11.65 | 9.83 |
| | 0.9 | 3 | 1.18 | 1.37 | 1.55 | 1.60 | 1.60 | 1.53 | 1.25 |
| | | 5 | 1.26 | 1.68 | 2.75 | 3.11 | 3.20 | 3.16 | 2.64 |
| | | 10 | 1.31 | 1.87 | 6.56 | 9.61 | 10.42 | 10.59 | 8.95 |
| | 0.8 | 3 | 1.16 | 1.33 | 1.48 | 1.51 | 1.51 | 1.45 | 1.19 |
| | | 5 | 1.25 | 1.65 | 2.57 | 2.86 | 2.94 | 2.90 | 2.43 |
| | | 10 | 1.31 | 1.86 | 6.14 | 8.71 | 9.40 | 9.54 | 8.09 |
| | 0.7 | 3 | 1.15 | 1.29 | 1.41 | 1.44 | 1.43 | 1.36 | 1.14 |
| | | 5 | 1.24 | 1.61 | 2.39 | 2.62 | 2.69 | 2.64 | 2.22 |
| | | 10 | 1.31 | 1.84 | 5.71 | 7.81 | 8.39 | 8.49 | 7.25 |
| | 0.6 | 3 | 1.13 | 1.24 | 1.34 | 1.36 | 1.35 | 1.29 | 1.09 |
| | | 5 | 1.22 | 1.56 | 2.20 | 2.38 | 2.43 | 2.39 | 2.04 |
| | | 10 | 1.30 | 1.82 | 5.24 | 6.92 | 7.36 | 7.45 | 6.42 |

**Theorem 2.** *Under model* (1.1) – (1.2) *with* $\sigma_\delta^2$ *known, the ARE's of* $\hat{\alpha}(\theta)$ *with respect to* $\hat{\alpha}$ *and of* $\hat{\beta}(\theta)$ *with respect to* $\hat{\beta}$ *are greater than or equal to* $(1-2\theta)^2$.

**Proof.** Since the distribution $H$ of $\varepsilon_i^*$ is symmetric about 0, it follows that

$$\frac{1}{2}\sigma_{\varepsilon^*}^2 = \int_0^\infty s^2 dH(s) \geq \int_0^{\eta_2} s^2 dH(s) + \eta_2^2\theta = \frac{\sigma^2(\theta,H)}{2}(1-2\theta)^2.$$

Consequently, we have $\sigma_{\varepsilon^*}^2/\sigma^2(\theta,H) \geq (1-2\theta)^2$. From (3.4) and Remark 1, the ARE of $\hat{\beta}(\theta)$ with respect to $\hat{\beta}$, denoted as $e_{t,u}$, is given by

$$e_{t,u} = \frac{c+\sigma_{\varepsilon^*}^2}{c+\sigma^2(\theta,H)} = \frac{\sigma_{\varepsilon^*}^2}{\sigma^2(\theta,H)} + \frac{[1-\sigma_{\varepsilon^*}^2/\sigma^2(\theta,H)]c}{\sigma^2(\theta,H)+c},$$

where $c = 2\beta^2(1-r)^2\sigma_X^2 > 0$. When $\sigma_{\varepsilon^*}^2 \leq \sigma^2(\theta,H)$, obviously $e_{t,u} \geq \sigma_{\varepsilon^*}^2/\sigma^2(\theta,H)$ $\geq (1-2\theta)^2$. On the other hand when $\sigma^2(\theta,H) < \sigma_{\varepsilon^*}^2$, we have

$$e_{t,u} \geq \frac{\sigma_{\varepsilon^*}^2}{\sigma^2(\theta,H)} + 1 - \frac{\sigma_{\varepsilon^*}^2}{\sigma^2(\theta,H)} = 1.$$

A similar proof can be established for the ARE of $\hat{\alpha}(\theta)$ with respect to $\hat{\alpha}$.

## 5. Summary

We have considered a method of defining trimmed estimation in the structural errors-in-variables model. This method is accomplished by first writing the errors-in-variables model as a standard linear regression model. Then by applying the trimmed least squares estimators proposed by Koenker and Bassett (1978), the corresponding trimmed estimators for the errors-in-variables model are defined. These trimmed estimators are more efficient than the traditional consistent estimators and should be recommended when the regression error (i.e. the error $\varepsilon_i$ in (1.2)) in the model has a heavy tailed distribution. On the whole, the trimmed estimators with trimmed proportions between 10% and 20% work well. However, a better idea is to find a data-dependent trimmed proportion and this will be pursued in some other work.

The approach used here can be easily generalized to the case where there is more than one covariate. It also can be applied to the case where some covariates are measured exactly and some are measured with errors.


## Acknowledgment

We thank the referees for their comments, which helped us greatly in improving the presentation.

## References

Anderson, T. W. (1984). Estimating linear statistical relationships. Ann. Statist. **12,** 1-45.

Cheng, C. L. and Van Ness, J. W. (1991). On estimating linear relationships when both variables are subject to errors. Technical report #206, Programs in Mathematical Science, University of Texas at Dallas, Texas.

Fuller, W. A. (1987). Measurement Error Models. John Wiley, New York.

Kendall, M. G. and Stuart, A. (1979). The Advanced Theory of Statistics, Vol. 2, 4th edition. Hafner, New York.

Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica **46,** 33-50.

Lehmann, E. L. (1983). Theory of Point Estimation. John Wiley, New York.

Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Assoc. **54,** 173-205.

Meketon, M. (1986). Least absolute value regression. AT&T Bell Laboratories, Holmdel, NJ 07733, U.S.A.

Moran, P. A. P. (1971). Estimating structural and functional relationships. J. Multivariate Anal. **1,** 232-255.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. J. Amer. Statist. Assoc. **75,** 828-838.

Stigler, S. M. (1977). Do robust estimators work with real data? Ann. Statist. **5,** 1055-1098.

Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan.

(Received August 1993; accepted November 1994)